R E P O R T  R E S U M E S

ED 019 000 EM 006 335

A COMPUTERIZED DETERMINATION OF THE READABILITY OF PROGRAMMED MATERIALS USING COMPLETE UNITS.

BY- SHAW, ROBERT A. JACOBSON, MILTON D.

EDRS PRICE MF-$0.25 HC-$0.72 16P.

DESCRIPTORS- *PROGRAMED MATERIALS, *READING DIFFICULTY, *EDUCATIONAL EXPERIMENTS, SCIENCE UNITS, GRADE 4, GRADE 5, GRADE 6,

THE INCREASED USE OF PROGRAMED MATERIALS CREATE A NEED TO ESTABLISH PROCEDURES FOR DETERMINING THE READING DIFFICULTY OF THE MATERIALS BEFORE EXTENSIVE FIELD TESTING BEGINS. AN EXPERIMENT WAS CARRIED OUT WITH 189 FOURTH GRADE, 271 FIFTH GRADE, AND 140 SIXTH GRADE STUDENTS. THE POPULATION OF MATERIALS FOR THIS STUDY CONSISTED OF NINE FOURTH GRADE PROGRAMED SCIENCE UNITS. THE DEPENDENT VARIABLE WAS DEFINED AS THE AVERAGE NUMBER OF ERRORS THAT EACH STUDENT MADE IN THREE SELECTED PROGRAMED SCIENCE UNITS. SEVENTEEN INDEPENDENT VARIABLES WERE CHOSEN AS TO REPRESENT FUNCTIONAL AND GRAMMATIC ASPECTS OF THE UNITS. REGRESSION ANALYSIS WAS UNDERTAKEN TO DETERMINE THE TOTAL AND RELATIVE INFLUENCE OF THE INDEPENDENT VARIABLES ON ERROR COUNT. THE ERROR COUNT WAS FOUND VALID AS A CRITERION FOR DETERMINING READING DIFFICULTY BECAUSE (1) CORRELATIONS WITH THE RESULTS OF THE UNIT TESTS WERE ALL SIGNIFICANT IN PREDICTED DIRECTION, AND (2) BECAUSE THREE INDEPENDENT VARIABLES WERE SIGNIFICANT AS PREDICTORS OF VARIATIONS IN ERROR COUNT. IN FUTURE STUDIES, IT IS RECOMMENDED THAT INDEPENDENT VARIABLES BE GIVEN AMPLE CONSIDERATION, AND THAT SAMPLE BE ENLARGED TO INCLUDE LARGER NUMBER OF UNITS. (OH)

ED019000

# A COMPUTERIZED DETERMINATION OF THE READABILITY OF PROGRAMMED MATERIALS USING COMPLETE UNITS

Robert A. Shaw
University of Connecticut

Milton D. Jacobson
University of Virginia

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

## INTRODUCTION

Determining the reading difficulty of books is not a recent development, but only in the last few years had the attempt been made to examine the readability of books by scientific procedures. Many types of materials have been examined but little work has been done with the readability of programmed materials. Yet indications are that programmed materials will be of increasing importance; therefore, there is a need to establish procedures for determining the reading difficulty of the materials before extensive field testing begins.

## STATEMENT OF THE PROBLEM

The purpose of this study was to automate a procedure which could be used to aid in determining the reading difficulty of selected programmed materials. The study involved: (1) the development of a regression equation which could be used to predict the reading difficulty of the given materials; (2) the selection of samples of various sizes and types to determine under which sampling procedure the best predictions of readability

EM 000 252

could be made in relation to the total sample; and (3) the evaluation of the sequence of the materials to determine if the assumptions of good sequential development were satisfied.

## TECHNIQUES AND PROCEDURES

In order to achieve the purposes of this study, the following steps had to be employed:

1. Independent variables appropriate for programmed materials were selected.
2. Computer programs were developed which count all of the independent variables automatically.
3. The criterion variable was defined as the average number of errors that each student made in the units.
4. Regression equations were determined for the criterion and independent variables.
5. Different sampling techniques were compared with the total sample effectiveness.

### Selection of Materials

The population of materials for this study consisted of nine fourth grade programmed science units. From the nine units three were selected to be included in the sample: (1) sound, (2) light, and (3) heat. The three units under investigation were constructed, linear programmed units which had been tested and revised.

## Selection of the Students

The programmed materials were used in twenty elementary schools in Central Virginia during the school year 1966-67. The population of students who studied science by means of the programmed science materials consisted of 140 sixth graders, 271 fifth graders, and 189 fourth graders.

## Determining the Variables

### Error Count for the Criterion Variable

To establish a criterion variable which could be obtained from the completed units and still be a measure of reading difficulty, an error count was made. The number of errors was obtained for each individual per sample of material, and the average number of errors was calculated. The error count was defined as the relative difficulty of a frame.

### The Independent Variables

From an analysis of the previous readability studies and the materials under investigation, the following variables were determined for each sample:

$X_1$: Average number of paragraphs per frame.

$X_2$: Average number of sentences per paragraph.

$X_3$: Average number of words per sentence.

$X_4$: Average number of letters per word.

$X_5$: Average number of simple sentences.

$X_6$: Average number of words which were outside Thorndike's list of 6000 words (measure of difficult words).

$X_7$: Average number of mathematical and scientific words.

$X_8$: Average number of mathematical and scientific numerals or symbols.

$X_9$: Per cent of frames that were response frames per sample.

$X_{10}$: Per cent of response frames that were structured response frames.

$X_{11}$: Per cent of response frames that were free response frames.

$X_{12}$: Per cent of frames that were non-response frames.

$X_{13}$: Per cent of responses requiring mathemtical or scientific words.

$X_{14}$: Average number of frames using same key word or phrase consecutively.

$X_{15}$: Average number of words in phrases per average number of words in phrases in succeeding frames (measure of redundancy).

$X_{16}$: Average number of disjoint frames.

$X_{17}$: Average number of review frames.

## The Regression Analysis

The counts of the variables for the total sample were used to obtain, by use of a computer, a regression equation which contained all of the variables. Using the F-ratio and the multiple R, a procedure was established for systematic deletion of variables. Regression equations were produced which contained the significant variables. A null hypothesis for each variable was either rejected or accepted at the 0.01 level of significance. A null hypothesis for the significance of the final equation was also determined, and an analysis of variance procedure was used to determine the status of this null hypothesis.

## Sampling Procedures

Once the proper deletions had been made three general sampling procedures were used to determine if and when a precision index could be reached for the independent variables. Two random samples were used, one with n = 50 pages and one with n = 97 pages. Using these samples, the established procedure of the regression analysis was repeated and equations produced. A similar technique was used with the two sequential samples. The first one-third of the total sample was used and then, the first two-thirds of the total sample was used. This sequential procedure was continued by taking the first one-third of each unit and putting these parts together. A cluster-sequential sample was the result. Two thirds of each unit were then grouped together for the final cluster-sequential sample. A null hypothesis with which to judge the significance of difference was stated for each different sampling procedure and the total sample, and between each sampling procedure.

## DESIGN AND ANALYSIS OF DATA

The number of errors made by the students who used the programmed science materials during the school year 1966-67 was obtained. The average number of errors made was used to measure the reading difficulty of the frames. The validation of the criterion variable produced negative correlation coefficients, as predicted, because the assumption had been made that the larger a student's technical vocabulary, the fewer errors he would make

in the program. The correlation coefficients were all significant at the 0.01 level. See Appendix I for a table of results.

## Total Sample

The total sample for this study consisted of three units: (1) Sound, (2) Light, and (3) Heat. Errors were counted for the students who completed these three units and took the unit tests.

## The Computer Program

The selection of variables, the counting of the variables, and the use of these variables have been composite problems in past studies of readability. This particular section was designed to eliminate the counting problem. A computer program was constructed into which could be read the entire units; listings and counts were the result. This program served the major purpose of using alphabetic characters and numerical characters simultaneously. Generally, the program involved sorting, locating, and accumulating the data. The data were the context and responses of the three programmed units. The counts obtained by using this program can be seen in Appendix II.

Besides this table the program produced two listings of words, one for context and one for the responses. All words in each list were alphabetized and the page and frame on that page in which the word appeared were listed.

## Basic Data

The data for the regression analysis consisted of the counts on the variables for the 188 pages of the three units. The data

were obtained by using the computer and the results of the basic program.

## Regression Analysis

To determine which independent variables played significant roles in predicting the criterion variable a stepwise multiple regression technique was used. In this process the variables were entered into the program in order of most significant to least significant.

Since a major objective of this investigation was to study the effects of the selected independent variables on the criterion variable, it was necessary to control the effects of the extraneous variables. A multiple regression technique was used in which full and restricted mathematical regression models were defined. The full model represents an attempt to express the dependent variable as a linear combination of all the other variables. The restricted models reflected the effects on the predictive information of factors considered in various forms. The purposes for the restricted models in this study were to test a null hypothesis for each variable and various combinations of variables.

## Hypotheses

The form of the hypotheses for the variables that were stated was as follows:

> $X_i$ (where $i = 1, 2, 3, \ldots, 17$) does not contribute to the regression equation to predict the reading difficulty as determined by average number of errors per sample ($a_i = 0$).

A null hypothesis was determined to test the significance of the final regression equation.

Regression Analysis Using Different Types of Samples

As a minor purpose of this study limited examinations were made as to size and type of sample which would enable valid predictions, and comparisons were made on means and standard deviations of the variables to determine those values which were the same to two significant numbers. The multiple R's were the only statistics subjected to testing. Appendix III contains the precision indexes for each variable to which each sample was compared. The full model and the basic restricted models remained the same as established with the total sample. Sequential sampling was used: (1) one-third of total and (2) two-thirds of the total. Cluster-sequential sampling was used: (1) one-third of each unit under investigation and (2) two-thirds of each unit under investigation. Random sampling was used: (1) a random sample of fifty pages and (2) a random sample of ninety-seven pages.

## FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

### Findings

From an analysis of the literature and an examination of the programmed materials seventeen independent variables were selected.

The error count, or dependent variable, was done using five measures of achievement and one IQ measure. All correlations were negative and significant at the 0.01 level of significance (see Appendix IV).

[illegible] hypothesis testing procedure, size, power, assumptions, and critical region must be specified. Suppose the null hypothesis

$$H_0: [illegible] = [illegible]$$

This equation must be true if hypothesis $H_0$ is true, otherwise the [illegible] must be unequal. The null hypothesis must be specified.

After the construction of the experiment equation and parameters for the null hypothesis and alternative hypothesis, the following procedure must be taken and the parameter null hypothesis must be tested:

1. Choose which is to be significant difference between the multiple comparisons chosen: (1) equivalent comparison, (2) unequivalent comparison, and (3) combination unequivalent comparison and the total comparison parameters.

2. Choose which is to be significant difference between the multiple comparisons of each comparison parameter: (1) equivalent comparison, (2) unequivalent comparison, and (3) combination-unequivalent comparison.

Hypothesis one was accomplished at the $0.01$ level of significance and hypothesis two was accepted at the same level.

## Conclusion

As a result of this study it was concluded that the equation above is the best prediction of the [illegible] of the multiple correlation.

The second assumption is best described as a combination comparison in determining [illegible] in parameters comparison: (1) the correlation between the level and the average number of observe were [illegible] and [illegible], and (2) the [illegible] correlation above. Each significance at the prediction of the observe and one characteristic of parameter [illegible] $\delta_p$ comparison a significance point is best [illegible].

It was concluded that under the conditions and assumptions of this study the three significant variables, $X_7$ (the average number of mathematical and scientific words (terms) per sample), $X_9$ (the per cent of frames that were response frames per sample), and $X_{17}$ (the average number of review frames per sample), were considered to be those variables that best predict the reading difficulty of the programmed science materials (see Appendix IV).

The materials did satisfy the assumptions of good sequential development: (1) as students progressed through the units their number of errors decreased, (2) the majority of the frames had a degree of overlap and the disjoint frames were, to a large degree, review frames, (3) frames were written from simple to complex as the number of mathematical and scientific terms did increase as the number of frames increased, and this was one of the significant variables.

It was further concluded that, with the exception of the total sample, the different sampling procedures used in this study did not serve to establish the same level of precision for all of the variables as with the total sample. Therefore, for this study, one partial sampling procedure could be rated as good as any other. However, some doubt was raised concerning small random samples as significant R's occurred between the random sample of fifty and the cluster-sequential samples.

From the above information it was concluded that, until different and more exact sampling procedures are developed, the total sample should be used, and the only feasible way to obtain a total sample is by use of a computer.

## Recommendations

For future studies concerning programmed materials, it is recommended that independent variables specific to programmed materials be given ample consideration, and that the computer program be extended so that, once the materials has been read in, the significant regression equation can be obtained without manual manipulation of variables.

It is also recommended that, as a future study, the sample be enlarged to included a larger number of units than was used in this study, and that the sampling procedures be broken down into smaller proportions to determine if and when a variable stabilized before the total sample.

From the conclusions of this study, it seems reasonable to recommend that programmed materials subjected to this type of analysis should be analyzed in terms of the concepts and skills to be taught by the program.

# BIBLIOGRAPHY

Bottenberg, Robert A. and Ward, Joseph H., Jr. Applied Multiple Linear Regression. Technical Documentary Report PRL-TDR-63-6. March, 1963.

Chall, Jeanne S. Readability, An Appraisal of Research and Application. Columbus, Ohio: The Ohio State University, 1948.

Chall, Jeanne S. "This Business of Readability," Educational Research Bulletin, XXVI, No. 1 (January 15, 1947), 1-13.

DeCecco, John P. Readings in Programmed Instruction: Educational Technology. New York: Holt, Rinehart, and Winston, 1964.

Flesch, Rudolph. "A New Readability Yardstick," Journal of Applied Psychology, XXXII, No. 3 (June, 1948), 221-233.

Hallinson, G., et. al. "Readability of Grade School Science Texts," School Review, LX (February, 1952), 94-98.

Schramm, Wilbur. The Research on Programmed Instruction. Washington, D.C.: U.S. Government Printing Office, 1964.

APPENDIX I

VALIDITY CORRELATION COEFFICIENTS FOR

ERROR COUNT

| Variable | Mean($\bar{X}$) | Standard Deviation(s) | Correlation Coefficient(r) | Significant (z with | = 0.01) |
|---|---|---|---|---|---|
| Error Count($Y_1$): | 42.76 | 44.29 | | | |
| STEP Test ($Y_3$): | 32.81 | 10.76 | -0.3055 | -7.75 | * |
| Unit Tests: | | | | | |
| Sound($Y_6$): | 23.59 | 5.22 | -0.2683 | -6.75 | * |
| Light($Y_7$): | 22.71 | 6.94 | -0.3223 | -8.00 | * |
| Heat($Y_8$): | 18.82 | 6.01 | -0.2144 | -5.25 | * |
| T-P Test ($Y_4$): | 113.09 | 17.53 | -0.1483 | -3.75 | * |
| IQ Test ($Y_2$): | 98.95 | 14.36 | -0.3925 | -9.75 | * |
| N = 559 | | | | | |

$Y_5$ was a corrected IQ score, but since not all students had this score it was not reported.

* Significant at the 0.01 level of significance.

# APPENDIX II

## VARIABLE COUNTS VIA STREAM PROCEDURE

| | |
|---|---|
| Total Number of letters | 79926 |
| Total Number of Words | 18994 |
| Total Number of Sentences | 1677 |
| Total Number of Paragraphs | 1012 |
| Total Number of Frames | 902 |
| Number of Response Frames | 813 |
| Number of Structured Response Frames | 808 |
| Number of Free Response Frames | 4 |
| Number of Non-Response Frames | 89 |
| Number of Review Frames | 75 |
| Number of Disjoint Frames | 149 |
| Number of Non-Disjoint Frames | 753 |

## APPENDIX III

PRECISION INDEXES OF THE VARIABLES IN FULL MODEL

(MEANS FROM TOTAL SAMPLE)

| Variable | Precision Indexes |
|---|---|
| $X_1$ | 1.23 |
| $X_2$ | 1.76 |
| $X_3$ | 11.95 |
| $X_4$ | 4.99 |
| $X_5$ | 3.59 |
| $X_6$ | 1.76 |
| $X_7$* | 3.86 |
| $X_8$ | 0.71 |
| $X_9$* | 82.77 |
| $X_{10}$ | 94.10 |
| $X_{11}$ | 2.18 |
| $X_{12}$ | 17.23 |
| $X_{13}$ | 1.71 |
| $X_{14}$ | 3.99 |
| $X_{15}$ | 8.64 |
| $X_{16}$ | 0.80 |
| $X_{17}$* | 0.41 |
| Y* | 0.22 |

*The significant variables.

APPENDIX IV

SUMMARY OF REGRESSION ANALYSIS

| | Variables | Model | R | F | for $df_1$ and $df_2$ | | Significance at 0.01 level hypothesis |
|---|---|---|---|---|---|---|---|
| (1) | All Variables | MF | 0.573 | | | | |
| (2) | $X_1, X_3, X_4, X_5, X_6, X_8, X_{12}, X_{13}, X_{14}, X_{15}$, and $X_{16}$ deleted | $MR_1$ | 0.561 | 0.3482 | 11 | 170 | acc. |
| (3) | All of #2 and $X_{11}$ deleted | $MR_2$ | 0.542 | 0.7726 | 12 | 170 | acc. |
| (4) | All of #3 and $X_2$ deleted | $MR_3$ | 0.524 | 1.0983 | 13 | 170 | acc. |
| (5) | All of #4 and $X_{10}$ deleted | $MR_4$ | 0.502 | 1.4642 | 14 | 170 | acc. |
| (6) | All of #5 and $X_7$ deleted | $MR_5$ | 0.451 | 2.2191 | 15 | 170 ** | rej. |
| (7) | All of #6 and $X_9$ deleted | $MR_6$ | 0.342 | 3.5417 | 16 | 170 ** | rej. |
| (8) | $X_{17}$ only deleted | $MR_7$ | 0.528 | 12.8308 | 1 | 170 ** | rej. |